

Evaluating “superiority”, “equivalence” and “non-inferiority” in clinical trials

Fatma Nesrin Turan,* Mustafa Şenocak†

From the *Department of Biostatistics, Trakya University, Edirne and the †Department of Biostatistics, Istanbul University, Istanbul, Turkey

Correspondence and reprint requests: Trakya University · Department of Biostatistics · Güllapoglu Campus · Edirne 22030 · Turkey · fnes-rinturan@yahoo.com.tr · Accepted for publication January 2007

Ann Saudi Med 2007; 27(4): 284-288

Clinical studies are usually performed with the aim of justifying that a new treatment approach is “superior” to the common standard approach (active control) with respect to benefits. In a general sense, this justification is carried out on the basis of the “null hypothesis significance test” with the P value based on this test used for justification. Today, new drugs differ so little from existing ones that factors such as cost and side effects affect the choice of therapy, when the bioavailability of treatment methods are found equivalent. Therefore, the aim of comparative clinical trials has extended beyond showing that a treatment is “superior” and now attempts to show that new treatments are “equal” and “non-inferior” to existing treatments. New approaches have become necessary since the classical null hypothesis approach is insufficient to justify the use of new agents, especially in cases of “equivalence” and “non-inferiority”. This new approach to justification makes use of the “clinical equivalence interval”, which determines the limits of the differences between specific endpoints that can be regarded as clinically “equal” to the value that was pre-specified based on studies of established therapies. It also makes use of the quantitative-based “confidence intervals” as the criteria for statistical justification. Many analyses can be done confidently when these tools are applied and the data are interpreted correctly.

Clinical studies are usually performed to reach conclusions about the superiority of one treatment over another. The concept of “superiority” might have different characteristics in various contexts. Generally, this justification is carried out using statistical significance tests to reject the “null hypothesis”. This brings up the point that failing to show superiority does not mean that there is equivalence.¹ Today, the aims of clinical research exceed simply showing that a treatment is “superior” to the other by clinical effect size. Clinical research also aims to address issues related to “equivalence” and “non-inferiority.” If there is no difference between two drugs that are equal in clinical effectiveness but not in other characteristics, such as side effects, costs and ease of use, then these other factors will result in selection of the best drug.²⁻⁵ Therefore, clinical research evaluations are designed to test:

- ✦The superiority of a new initiative (drug or treatment) to an available initiative,
- ✦The non-inferiority of a new initiative to an

available initiative,

- ✦The equivalence of a new initiative to an available initiative.

Irving et al stated that when designing a non-inferiority/equivalence trial the investigator intends to show efficacy by demonstrating that a new treatment is as good as or not worse than a known effective treatment by a small predefined margin.⁶ However, the classical null hypothesis and the P evaluation on the basis of the classical null hypothesis are not enough to justify “equivalence” and “non-inferiority”.^{2,7,8} New concepts and approaches have been designed to provide reliable justifications in the selection of the most effective therapy. In our review, we will briefly explain these approaches.

Approaches to justification

The typical approach to evaluate the results of a clinical study related to two (or more) comparative sets of quantitative or qualitative data is based on a P value obtained through a statistical analysis. The inadequacy of

P values stems partly from theoretical questions about their meaning and interpretation, and also from the fact that they do not in themselves convey adequate information about the size or direction of the effect or the range of possible outcomes.^{2,8} Clinical features of the study might result in falsely evaluating differences, which have no clinical importance, as reflecting superiority in the acceptance of the null hypothesis. This point will especially be important in “justifications excluding superiority”. The criterion for equivalence determines the difference that will be regarded as “equivalent” to the mean value, denominated by an interval called “ Δ ”, which is determined by the clinician and is predicted-expected-known in a specific clinical situation.⁵ For example, if a clinician regards a decrease of only 30 mm Hg as a clinically important difference as opposed to a decrease of 10 mm Hg in the systolic blood pressure, which is predicted as 165 mm Hg, then the “interval for equivalence” is 30 mm Hg. The statistical analysis of comparative clinical trials is generally based on “confidence intervals”,⁹ and the P value obtained is based on values obtained by previous research. Confidence intervals provide a more reliable comparative measurement of the real difference, which is more interesting and important than a P value. Also, confidence intervals give more detailed information about superiority, equivalence, and non-inferiority than the P value. Therefore, the upper and lower values

of the confidence interval can be interpreted as the efficiency of the treatment with a high level of certainty. Thus, the confidence interval demonstrates the size of the effect as well as the lowest and the highest margins of the estimate in addition to evaluating the null hypothesis.^{2,8}

Factors other than the statistical analysis are important to consider in judging “equivalence” and “non-inferiority”. In this context, the methodological principles that should be used to compare a new initiative with common control initiatives^{2,3} are to ensure that equivalence and noninferiority criteria are predefined, to determine the eligible conditions for the study, in which the efficacy of the common control initiative was established, to apply both initiatives under eligible conditions, and to allow sample size estimates to be based on the correct power calculations. Thus, the clinical study, which will include a comparative justification, must have three basic statistical elements apart from the methodological characteristics, whatever its interpretation: the P value for the comparison, confidence intervals, and clinical equivalence intervals.

Justifications for Superiority

The justification for superiority is intended to prove that a new experimental drug is superior to the common control drug. The classical approach is the statis-

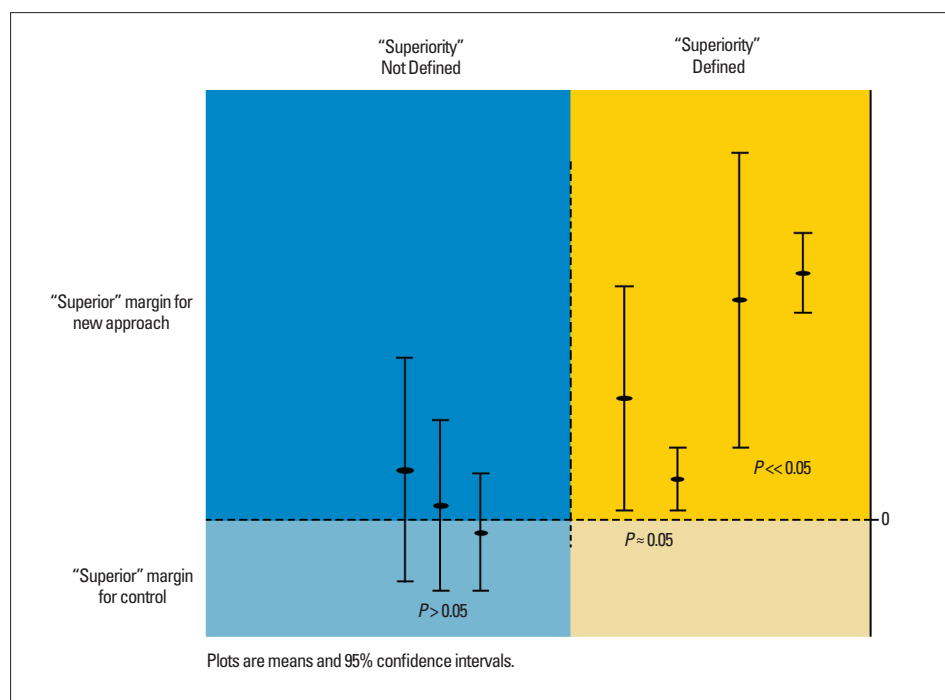


Figure 1. The statistical figuration of the evaluation for "superiority" using the approach of confidence intervals.

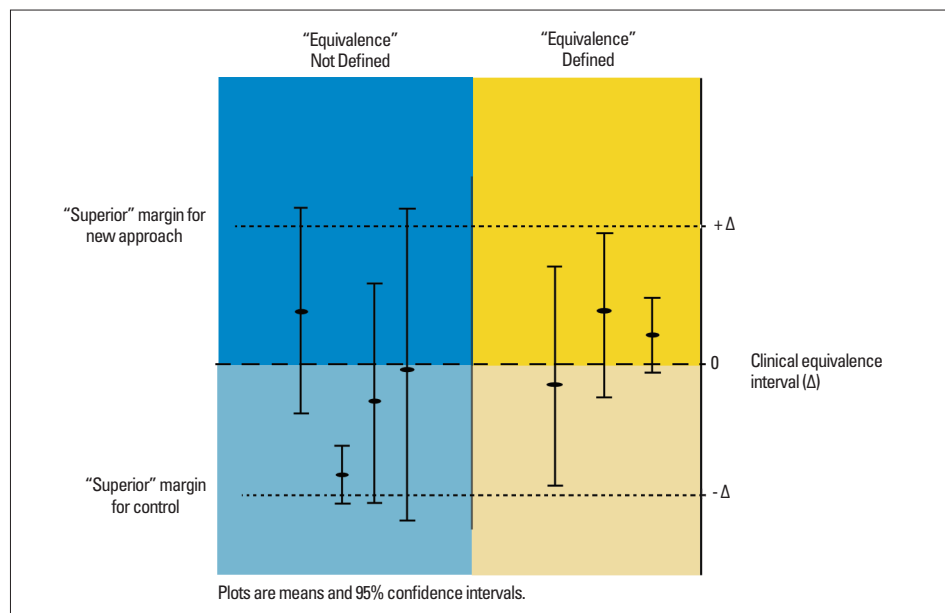


Figure 2. Confidence interval approach and evaluation of the equivalence with “ Δ ”.

tical justification that shows no difference between the clinical effects, including a declaration that the null hypothesis is met. First, once this difference is determined as not equal to “0” through statistical analysis, then the size of the difference to detect whether the effect is clinically adequate or not is estimated. The “real” value for the difference between treatment effects lies within the boundaries of the confidence interval (Figure 1). The “null hypothesis” (no difference between clinical effects) is rejected if the upper bound of the confidence interval for the difference between the test treatment and control is lower than the specified margin.² Thus, the two situations given below become equivalent to each other:

- ✦ The bilateral confidence interval of 95% for when the difference between the means includes “0”
- ✦ Two means are different from each other bilaterally when considered at the 5% level statistically.⁵

Naturally, the significance level can be regarded as different from 5%. If the “clinical equivalence interval” has been used to predict the number of samples at the beginning of the study, the result will be reliable. Yet, the clinician has the right to judge a difference seeming significant within his own criteria.

Justifications for Equivalence

When two drugs are absolutely equivalent it is not possible to find an exact equivalence of the predictive means (or rates) due to the biological variability of the phe-

nomena.⁹ Therefore, trials for equivalence are designed to prove the absence of a significant difference between the drugs. Temple et al² reported that such drugs would not show superior efficacy in active-control trials, yet equivalence to the active control with respect to efficacy would not have been informative.¹⁰ The evaluation starts with the prediction of the greatest acceptable “clinically identical” difference, and the “clinical equivalence interval”. Thus, it is accepted that the differences that exceed this level by occurring more or less than the exact amount are important clinically, and therefore, the clinical equivalence margin is chosen by identifying the clinically acceptable and the greatest difference in the justification for equivalence. If two drugs are equivalent to each other, then the bilateral confidence interval of 95% for the difference of effect between the drugs must cover the area between $-\Delta$ and $+\Delta$ (Figure 2). In this framework, there are equivalence margins, which can be asymmetrical with respect to zero, and they are determined by the clinician. In studies on bioequivalence, the confidence interval of 90% is accepted as a standard in the evaluation of the two drugs to find out whether their pharmacokinetic criteria means are equivalent to each other or not.⁵

Justifications for Non-inferiority

A noninferiority trial refers to a study in which the primary objective is to evaluate whether the new treatment is not inferior to or as effective as the standard therapy for a particular end point. It is not important

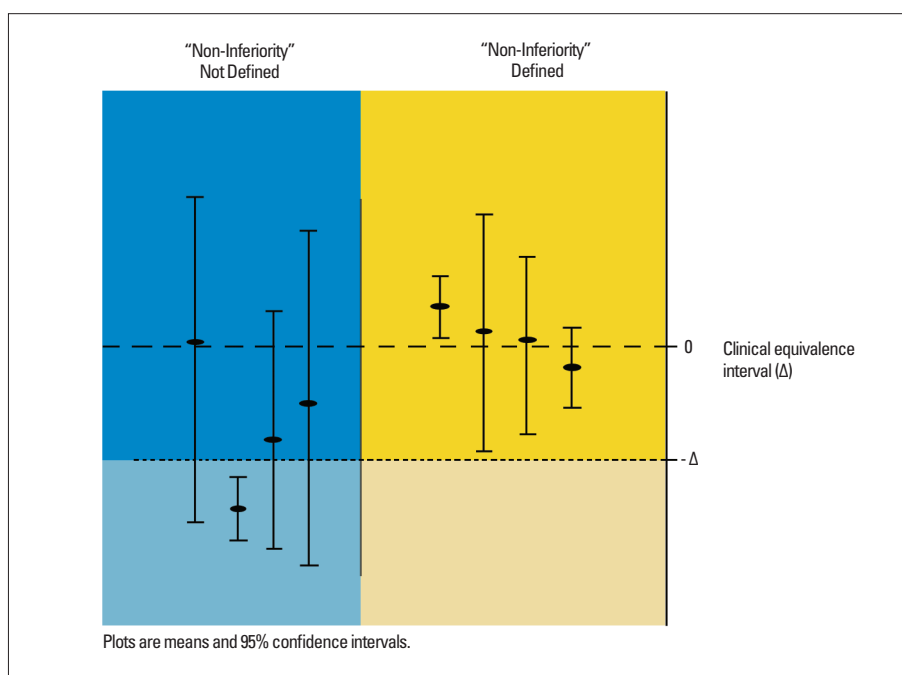


Figure 3. Confidence interval approach and evaluation of the non-inferiority with “ Δ ”.

to establish that the new drug is more effective or has similar effects, only that it is adequate to determine that it is not inferior to the drug to be compared. At this point, the evaluation starts with the prediction of the “clinical equivalence interval (Δ)” and continues by determining the bilateral confidence interval (90% or 95%). However, in this context, only a one-way probability of the difference is investigated. Thus, the confidence interval must extend to the “ $-\Delta$ ” side (Figure 3). The application of confidence intervals in the one-way justification meets the classical null test hypothesis that the difference of treatment effect that is equal to a lower margin of equivalence contrary to the alternative of a difference of treatment that is higher than the lower margin of equivalence level.¹² “Justifications for non-inferiority” might be designed in the same way as justifications for equivalence and cause misinterpretations, especially when the difference in treatment effect mentioned formerly is disregarded.^{5,9} In this context, the margin chosen for the justification for non-inferiority cannot be greater than the size of the lowest effect that the control drug is expected to have. The identification of the margin is based on both statistical inference and clinical justifications, and thus, it must reflect the uncertainties within the proof on which the choice is based and it must also protect its place appropriately. If these analyses are carried out properly, the finding that the confidence interval for the difference between

the common standard approach (active control) and the new drug excludes the margin, which has been chosen appropriately, might guarantee that the drug has a greater effect than zero. The margin, which has been chosen through projecting the size of the clinically acceptable effect, will probably be lower than the amount proposed with the expected lowest size of the effect of the standard approach.¹³

Conclusion

The chance of finding new treatment methods, techniques, and drugs that are superior to existing ones is so little that the differences between drugs for the same treatment are starting to be based on secondary characteristics such as price, side effects, and ease to use. However, differences based on these factors can only be made after comparing the efficiency of the treatment. Therefore, the aim of many studies is not only justifying that a treatment is superior to another, but also justifying that two treatments are equivalent to each other or that one of them is non-inferior to the other. If the evaluations are carried out on the basis of just simple statistical test results, inadequate or false result justifications might be obtained. It is very important to make the hypotheses with their characteristics clear in this kind of scientific study, as in other scientific research fields. The second important topic is the “clinical equivalence interval”, and this concept is mostly identified

through sizes of the intervals, which are determined by the clinician. The confidence interval for the difference between the comparative effects of the treatment provides the correct interpretation when it is justified with the other two concepts. The *P* level, which can be reached with classical statistical justifications, has

only a certain importance, and it may cause inadequate or even false justifications when it has been evaluated without considering the other concepts that have been mentioned. It is vitally important to justify clinical studies by examining them within the framework of the concepts.

REFERENCES

1. Gomberg-Maitland M, Frison L, Halperin JL. Active-control clinical trials to establish equivalence or noninferiority: Methodological and statistical concepts linked to quality. *Am J Med.* 2003; 116:398-403.
2. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. *Annals of International Medicine.* 2000;133(6):455-63.
3. McAlister FA, Sackett DL. Active-control equivalence trials and antihypertensive agents. *Am J Med.* 2001;111:553-8.
4. Committee for Proprietary Medicinal Products (CPMP): Points to Consider on the Choice of Non-inferiority Margin. London, 2004 February 26. CPMP/EWP/2158/99 draft.
5. Committee for Proprietary Medicinal Products (CPMP): Points to Consider on Switching Between Superiority and Non-inferiority. London, 2000 July 27. CPMP/EWP/482/99.
6. Irving KH, Toshihiko M. Design issues in non-inferiority/equivalence trials. *Drug Information Journal.* 1999; 33:1205-18.
7. Streiner DL. Research methods in psychiatry. *Can J Psychiatry.* 2003;48:756-61.
8. Huson L. Statistical assessment of superiority, equivalence and non-inferiority in clinical trials. *OR Focus.* 2001;12(5):31-4.
9. Miller JA, Burke V. Relationship between sample size and the definition of equivalence in non-inferiority drug studies. *J of Clinical Pharmacy & Therapeutics.* 2002;27(5):329-33.
10. Ellenberg SS, Temple R. Placebo-controlled trials and active-control trials in the evaluation of new treatments. *Annals of International Medicine.* 2000;133(6):464-70.
11. Chung B, Zee Y. Planned equivalence or non-inferiority trials versus unplanned noninferiority claims: Are they equal? *Journal of Clinical Oncology.* 2006;24(7):1026-1028.
12. ICH Topic E9 Statistical Principles for Clinical Trials. Step 4, Consensus Guideline, 5 February 1998. The European Agency for The Evaluation of Medicinal Products Human Medicines Evaluation Unit. London, 1998 March 18. CPMP/ICH/363/96.
13. ICH Guideline E10: Choice of Control Groups in Clinical Trials (Step 2). http://www.hc-sc.gc.ca/hpfb-dgpsa/tpd-dpt/e10_e.html.